用微博中个体主义-集体主义词语表达预测玩家对单机或网络游戏的偏好*

张伟斌[1,3] 王佳舟[2,3] 萨切尔[2,3] 王歆睿[2,3] 代启隆[2,3] 朱廷劭[1,3]

([1] 中国科学院心理研究所行为科学重点实验室，北京 100101)

([2] 中国科学院心理研究所心理健康重点实验室，北京 100101)

([3] 中国科学院大学心理学系，北京 100049)

摘　要：玩家对不同类型游戏的偏好会受到其自身特征的影响。游戏玩家数量，决定了单机游戏模式自主性更高，而网络游戏模式集体性更高。考虑到个体主义倾向个体强调独立，而集体主义个体强调协作。我们认为玩家个体主义-集体主义倾向，可能会影响其对于单机游戏或网络游戏的偏好。本研究利用微博大数据，使用文本分析的方法，旨在探讨单机游戏玩家及网络游戏玩家微博中个体主义-集体主义倾向词语及词类表达是否存在差异，并希望利用这些特征预测玩家对单机游戏或网络游戏的偏好。结果发现，单机游戏玩家微博中个体主义词类词语出现频率更高，而网络游戏玩家微博中集体主义词类词语出现频率更高。利用机器学习方法，仅用个体主义-集体主义词语表达做自变量，可预测玩家类型，但准确率较低。本研究为利用微博大数据识别用户对游戏的偏好提供了初步证据，具有一定的应用价值。

关键词：个体主义-集体主义；单机游戏；网络游戏；玩家类型；文本分析

1 引言

最近十几年来中国游戏玩家数量飞速上涨，调查显示 2015 年游戏玩家总数已超过 5.3 亿人（王艳, 2016）。游戏开发商不断推行出多种多样的游戏，以满足市场中的不同需求。从玩家数量角度来划分，游戏的种类可大致分为单机游戏及网络游戏，在本研究中，单机游戏指只有一个玩家参与的游戏，而网络游戏指有多名玩家同时在线并且可以发生互动的游戏。

日常生活中，我们观察到单机游戏玩家和网络游戏玩家似乎是不同的人群。换句话说，一部分玩家更倾向于选择单机游戏，而另一部分玩家更倾向于

* 通讯作者：朱廷劭，Email：tszhu@psych.ac.cn

选择网络游戏。为什么人们会偏好不同类型的游戏？前人研究证实，玩家对游戏的选择受到其个体特征的影响，玩家会选择与自身特性更匹配的游戏（Chory & Goodboy, 2011; Fang & Zhu, 2011）。

单机游戏及网络游戏最本质的不同之处在于能否获取其他玩家的支持。对于单机游戏来说，玩家需依靠自身完成任务，因此独立性更高；而对于网络游戏来说，玩家间可以进行互动并联合完成任务，因此集体性更高。因此，偏好单机游戏的玩家可能更喜欢独立完成任务，而偏好网络游戏玩家可能更喜欢在集体中完成任务。两类玩家在这一特征上差异的或许可以用心理学领域中个体主义-集体主义倾向的概念来概括。研究中普遍认为，个体主义倾向个体，更关注自身的自主性和独特性；而集体主义倾向个体，更关注群体的利益和成员间的协作（苏红，任孝鹏, 2014; Oyserman, Coon, & Kemmelmeier, 2002; Hofstede, 1991）。因此，对于偏好单机游戏的玩家，他们个体主义倾向可能更高；而对于偏好网络游戏的玩家，他们集体主义倾向可能更高。换句话说，我们认为，玩家个体主义-集体主义倾向，可能会影响其对于单机游戏或网络游戏的偏好。

前人对于个体主义-集体主义倾向测量主要采用自我报告（Triandis & Gelfand, 1998），或者实验任务的方式（Talhelm et al., 2014）测得。然而，这些研究的结果可能受到实验者效应、社会赞许性等影响。最近一些研究，利用大数据的方式，描绘出谷歌图书数据库中特定词语出现频率随年份迁移的变化趋势，实现了国家层面个体主义-集体主义倾向变化趋势的量化评估（Greenfield, 2013; Zeng & Greenfield, 2015）。这种研究方式为个体主义-集体主义倾向的测量提供了新的思路。因此，本研究中，我们采用了类似的方式，从玩家大量的微博文本中，计算出个体主义-集体主义相关词语及词类的词频，最终希望通过利用这些词频特征预测玩家更偏好单机游戏还是网络游戏。

本研究共有两个目的，第一个目的是，探究单机游戏及网络游戏玩家微博表达中，个体主义及集体主义关键词及词类表达的差异。我们认为单机游戏有更多个体主义的词语表达，网络游戏玩家有更多集体主义的词语表达。具体假设如下：H1）网络游戏玩家，表达出更多“我们”；单机游戏玩家，表达出更多“我”；H2）网络游戏玩家，表达出更多集体主义倾向词类词语如“帮助，友情”等；单机游戏玩家，表达出更多个体主义倾向词类词语如“自主，个性”等。

第二个目的，也是更主要的目的是，使用机器学习的方法，利用个体主义及集体主义关键词及词类表达预测玩家单机游戏及网络游戏的选择。

2 方法

2.1 游戏筛选

在本研究中，单机游戏玩家特指“太吾绘卷”（单人策略角色扮演游戏，SRPG）游戏玩家。网络游戏玩家特指“逆水寒”（大型多人在线角色扮演游戏，MMORPG）游戏玩家。

对单机游戏及网络游戏的选取有以下考虑：1）我们参考了一些网站中的热门游戏排名，发现在中国，在当前时段，这两个游戏属于热门游戏；2）“太吾绘卷”和“逆水寒”都是国产电脑端游戏，都在 18 年 9 月左右发行或上线，都是古风背景，具有一定的可比性；3）“太吾绘卷”只有单人模式，“逆水寒”只有网络模式，一些热门游戏如“我的世界”，既有单机模式，也有网络联机模式，而本研究希望把“单人，多人在线”两种玩家区分，因此不考虑纳入这些游戏。

2.2 玩家微博数据获取

在微博上分别搜索以下两个关键词：1）对单机（“太吾绘卷”）玩家来说，关键词为“#太吾绘卷#”；2）对网游（“逆水寒”）玩家来说，关键词为“#逆水寒#”。利用爬虫的方式把提到该关键词的用户信息和微博文本抓下来，即可得到单机玩家及网游玩家数据。

在微博中定位玩家时，主要有以下两点考虑：1）用户在微博上提到某一游戏，如微博文本中出现“太吾绘卷”，并不一定代表该用户是太吾绘卷的玩家。但是如果用户微博文本中出现的是“#太吾绘卷#”，则说名该用户参与了该游戏的话题讨论，而参与话题讨论的用户，据我们不完全观察推断其很大概率是该游戏的玩家。因此在搜索玩家时，我们使用的关键词是“#太吾绘卷#”及“#逆水寒#”；2）对于搜到的用户账号，会去掉其中的加 V 账号，即主播、俱乐部、游戏制作运营方官方微博账号。因为这些账号属于公众媒体，其微博表达方式和大众玩家会不太一样，并可能会污染研究结果，因此予以剔除。

最终我们共得到 124 名单机游戏（“太吾卷绘”）玩家的微博数据，及 1000 名网络游戏（“逆水寒”）玩家的微博数据。

关于为什么两类玩家数据量差异这么大（约 10 倍），我们认为可能的解释是，网游（“逆水寒”）的官方微博经常会有一些转发抽点券的活动，因此参与该游戏话题的网游玩家数量就会更多。而对于单机游戏（“太吾绘卷”），没有什么官方活动，仅凭玩家自愿参与游戏话题讨论，因此相对数量就会更少。这个事实说明，如果想要用游戏相关的表达定位游戏玩家，特别是单机游戏玩家是较难的，因为微博中对游戏进行表达的人相对是少数。因此，间接反映了本研究意义——如果可以利用用户日常表达预测其对游戏的偏好，则可以应用到更广泛的用户群体中。

2.3 玩家样本选择

我们在 1000 个网游玩家中随机抽取 120 名玩家，作为研究中分析的网游玩家样本。而单机玩家样本为 124 名玩家。

我们之所以抽取部分网游玩家进行分析，是因为考虑到原始单机玩家和网游玩家样本数量差距较大，可能造成之后差异性检验的偏差。因此我们选择从中抽出一部分样本，与单机玩家样本量进行匹配。

所有网游玩家样本的微博注册年的平均值为 2010.45（SD=0.62），其中男性占比 26.3%，与抽出的网游玩家样本差异不大（微博注册年：M=2010.38，SD=0.63；男性占比：27.50%），因此可以认为研究中分析的网游玩家样本可以代表整体网游玩家样本。

2.4 特征值提取

对微博数据的分析，即对微博中抽取特征值的分析。我们从玩家微博中提取的特征值可分为两大部分：1）用户特征，即从用户状态中提取出的特征，包括每一个用户的性别、微博注册年、粉丝数、互粉数、关注数、点赞数；2）微博表达特征，即从用户微博文本中提取出的特征，包括每一个用户总微博数量、原创微博数量、原创微博总字数，“我”词频、“我们”词频、个体主义词典中词语词频、集体主义词典中词语词频。对特征值的提取使用 python 程序实现。

我们之所以选择这些特征值，主要是基于研究目的和假设。我们特别关注的是，1）个体主义-集体主义词语表达频率；2）个体主义-集体主义词类表达，即个体主义词典中的词语词频、集体主义词典中的词语词频，是否会在单

机玩家和网游玩家群体中出现差异。对微博文本特征值的提取，仅使用用户原创微博，因为原创微博更能反应用户本身特征。

参考前人文献，选择“我”“我们”分别代表个体和集体主义倾向词语（Dwell et al., 2011），选择“不同”“独立”等词形成个体主义词典，选择“一起”“家庭”等词形成集体主义词典，词典中包含的其他词见表 1（Xu & Hamamura, 2014; Twenge, Campbell, & Gentile, 2012; 包寒吴霜 等，2018）。

表 1 个体主义词典及集体主义词典词列表

词典名称	词列表
个体主义词典	['不同', '成就', '独立', '能力', '小资', '不一样', '民主', '权利', '开放', '私人', '个性', '权力', '多样化', '特立独行', '自立', '多元化', '独自', '竞争', '单独', '自主', '自由', '独处', '创新', '个体', '独特', '自己']
集体主义词典	['家人', '传统', '亲情', '协作', '家庭', '亲人', '共享', '谦虚', '谦逊', '群众', '义务', '抱团', '谦让', '妥协', '共同', '团队', '共度', '亲近', '服从', '归属', '和谐', '友情', '无私', '分享', '协同', '合作', '帮助', '集体', '顺从', '团体', '互助', '奉献', '团结', '一起', '友谊']

2.5 分析

（1）差异性检验

为了探究两类玩家用户特征及微博表达特征的差异，使用 python 中的计算函数，对单机玩家及网游玩家微博数据中提取的特征值（如微博中“我”出现的词频）进行差异性检验。因为所有特征值的分布均不符合正态分布，因此对所有变量进行曼-惠特尼 U 检验。

（2）预测玩家类型

利用机器学习中 SVM 方式，抽取玩家微博表达中若干特征值（如性别）作为自变量，来预测玩家类别（因变量）。对于自变量的选择有两点考虑，1）根据差异性检验的结果，挑选在两类玩家间有差异的特征值，2）选取研究比较关注的特征值。其中，考虑到词频特征值的原始数值过小，使用词频特征值作

为自变量时，将其原始数据乘以 100 后再使用。对因变量的处理为，把单机玩家编码为“0”，把网游玩家编码为“1”。

随机抽取所有样本（样本中即包括单机玩家及网游玩家，共 244 名玩家）中的 80%作为训练集，剩余 20%作为测试集。输入训练集的自变量和因变量，利用 SVM 方式进行学习，得到预测模型。之后利用该模型，输入测试集的自变量，预测其因变量（即玩家类型）。最后，使用准确预测数量除以总预测数量得到模型准确率，作为模型评估的标准。准确率越高，代表该模型对玩家分类的预测更好。

考虑到研究中样本量较小，使用单次抽样对应的模型预测，可能会产生较大偏差。因此我们共随机抽取 500 个训练集，形成 500 个模型，并计算每个模型的准确率，然后取 500 个准确率的平均值作为最终模型评估指标。

3 结果

3.1 差异性检验结果

两类玩家微博数据中提取的特征值的描述性统计结果及差异性检验结果如表 2 所示。

对于用户特征来说，我们发现单机玩家中男性玩家稍多（占 54.84%），而网游玩家中男性玩家非常少（占 27.50%）；单机玩家的互粉数（$p<0.01$）和关注数（$p<0.05$）显著高于网游玩家；而两类玩家微博注册年均为 2010 年左右，没有显著差异。

对于微博表达特征来说，我们发现两类玩家的总微博数没有显著差异。但是单机玩家原创微博数（$p<0.01$）和原创微博总字数（$p<0.05$）显著高于网游玩家。对个体主义-集体主义词语表达来说，网游玩家微博中出现的“我”（$p<0.01$）“我们”（$p<0.05$）两个词的词频更高。对个体主义-集体主义词类表达来说，单机玩家微博中表达个体主义词典中词语的词频更高（$p<0.05$），而网游玩家微博中表达集体主义词典中词语的词频更高（$p<0.05$）。

表 2 两类玩家特征值描述统计及差异检验结果

	单机（“太吾卷绘”）玩家		网游（“逆水寒”）玩家		
特征值	M	SD	M	SD	p

用户特征					
微博注册年	2010.38	0.63	2010.43	0.64	0.29
性别	1.45	0.50	1.73	0.45	0.00^{**}
点赞数	1144.98	3718.99	511.78	917.37	0.09
粉丝数	1826.28	6573.52	1289.84	3156.41	0.34
互粉数	175.54	142.72	129.69	115.02	0.00^{**}
关注数	523.53	396.97	451.52	313.22	0.04^{*}
微博表达特征					
总微博数	6920.56	3282.46	7059.78	5220.41	0.17
原创微博数	1830.73	1142.83	1759.09	2575.34	0.00^{**}
原创微博总字数	56103.65	37721.78	53537.48	53601.62	0.02^{*}
“我”词频[#]	184.87	57.03	205.26	60.50	0.00^{**}
“我们”词频[#]	18.77	12.35	22.45	15.62	0.02^{*}
个体主义词典词频[#]	47.97	22.79	43.58	20.20	0.03^{*}
集体主义词典词频[#]	49.47	39.54	53.56	35.73	0.02^{*}

备注：

性别：男性=1，女性=2。其中单机玩家中，男性占 54.84%；网游玩家中，男性占 27.50%。

个体主义词典：包括‘不一样’等词，具体词典见文中方法部分。

集体主义词典：包括‘一起’等词，具体词典见文中方法部分。

p 值由对两类游戏玩家某个特征值进行曼-惠特尼 U 检验得到。

* $p<0.05$，** $p<0.01$。

[#]为方便呈现，表中所示特征值的 M 及 SD 为原始数值的 10000 倍。

3.2 预测玩家类型结果

利用机器学习中 SVM 方式，抽取微博数据中的若干特征值作为自变量，使用玩家类型作为因变量，进行模型学习，并预测测试集中玩家类型，结果如表 3 所示。可以看到，仅使用性别作为自变量，500 个模型预测的平均准确率达到 63.43%。仅使用个体主义-集体主义表达集，即使用“我”“我们”词频，个体

主义词典词频，集体主义词典词频四个特征值作为自变量，模型预测的平均准确率达到 55.13%。但是使用性别，和个体主义-集体主义表达集，即使用五个特征值作为自变量，模型预测的平均准确率为 63.41%，较只有性别做自变量的预测模型并没有更高的提升。

表 3 用微博中特征值预测玩家类型的准确率

	准确率	
特征值（自变量）	M	SD
性别	63.43%	6.26%
个体主义-集体主义表达集	55.13%	6.55%
性别及个体-集体主义表达集	63.41%	5.86%

备注：

个体主义-集体主义表达集：其中共包括 4 个特征值，“我”词频、“我们”词频、个体主义词典词频、集体主义词典词频。

性别及个体-集体主义表达集：共包括 5 个特征值。

准确率=正确预测数/总预测数。准确率的 M 及 SD 由 500 次随机抽样预测结果得到。

4 讨论

在本研究中，我们从 124 名单机游戏（“太吾绘卷”）玩家及 120 名网络游戏（“逆水寒”）玩家的微博数据中提取出个体主义-集体主义词语表达频率、个体主义-集体主义词类表达频率等特征值。然后对两类玩家的特征值进行差异性检验。最终使用机器学习的方法，利用特征值完成对玩家游戏偏好的分类。结果显示，网游玩家在微博中会表达出更多集体主义词语，以及集体主义词类词语；而单机游戏玩家会表达出更多个体主义词类词语。使用微博中个体主义-集体主义表达做自变量建模，对玩家游戏偏好的分类的准确率，为 55%左右。但是使用性别做自变量建模，对玩家游戏偏好的分类准确率更高，为 63%左右。

本研究的第一个目的是，探究单机游戏及网络游戏玩家微博表达中，个体主义及集体主义关键词及词类表达的差异。我们认为单机游戏有更多个体主义

的词语表达，网络游戏玩家有更多集体主义的词语表达。我们的研究结果基本证实了研究假设，发现网络游戏玩家，在微博中表达出更多“我们”，及更多集体主义倾向词语如“帮助，友情”等；而单机游戏玩家，在微博中表达出更多个体主义倾向词语如“自主，个性”等。同时，结果发现单机游戏玩家原创微博数量，原创微博字数也多于网络游戏玩家。更喜欢在公众场所发表自己的观点，被认为是一种表达自身独特性的行为（Ho & Dempsey, 2010）。因此，这些原创微博数量特征同样说明，单机游戏玩家具有更高的个体主义倾高。然而，在我们的结果中发现了一个和预期假设不太相符的结果，即网络游戏玩家也会在微博中表达出更多“我”。对此，一个可能的解释是，单机游戏玩家在表达自我时可能不经常使用主语，比如“今天吃了夜宵真开心”这样缺乏主语的表达在偏口语化的微博表达中是可能出现的。综合以上结果，与 Chory 和 Goodboy（2011）研究一致，本研究再次证实了游戏玩家个体特质会影响到其对于游戏的选择，且玩家会倾向于选择和自身特性更为匹配的游戏。具体来说，研究发现个体主义倾向者偏好单机游戏，而集体主义倾向者偏好网络游戏。

本研究的第二个目的，也是更主要的目的是，使用机器学习的方法，利用个体主义及集体主义词语及词类表达预测玩家单机游戏及网络游戏的选择。结果发现使用个体主义-集体主义文本表达的方式训练出的模型可以预测玩家类型，但是其准确率仅略高于随机水平。可能是因为即便是同一类玩家，个体间差异也是非常大的，特征值较大的标准差也间接反映出这个事实。需要提到的一点是，目前用来训练模型样本量非常小，如果增加样本量，或许可以提高预测精度。同时，预测样本量也非常小，如果以目前的模型预测上万名没有目标值的微博用户，并得到他们对于单机游戏或者网游的偏好，尽管准确率只达到60%，但是成功预测的数量也会比较可观。除此之外，本研究中结果发现，单机游戏和网络游戏玩家群体性别差异非常大，因此之后的研究中如果要讨论玩家类型，性别是一个应该考虑在内的变量。总而言之，虽然目前的模型还可以进一步完善，但已经具备一定的应用价值。

综上所述，本研究发现微博表达中个体主义倾向高的玩家，更偏爱单机游戏，而微博表达中集体主义倾向高的玩家，更偏好网络游戏。利用微博表达中的个体主义-集体主义词语，及玩家性别可预测玩家对单机或网络游戏的偏好，

但是预测准确率还需进一步提高。本研究通过不同类型玩家的分析，实现游戏玩家类型的预测，可应用于游戏广告的投放，具有一定的商业应用价值。

参考文献

Chory, R. M., & Goodboy, A. K. (2011). Is basic personality related to violent and non-violent video game play and preferences?. *Cyberpsychology, Behavior, and Social Networking*, *14*(4), 191-198.

DeWall, C. N., Pond Jr, R. S., Campbell, W. K., & Twenge, J. M. (2011). Tuning in to psychological change: Linguistic markers of psychological traits and emotions over time in popular US song lyrics. *Psychology of Aesthetics, Creativity, and the Arts*, *5*(3), 200.

Fang, X., & Zhu, M. (2011, July). Extraversion and computer game play: who plays what games?. In *International Conference on Human-Computer Interaction* (pp. 659-667). Springer, Berlin, Heidelberg.

Greenfield, P. M. (2013). The changing psychology of culture from 1800 through 2000. *Psychological science*, *24*(9), 1722-1731.

Ho, J. Y., & Dempsey, M. (2010). Viral marketing: Motivations to forward online content. *Journal of Business research*, *63*(9-10), 1000-1006.

Hofstede, G. (1991). Organizations and cultures: Software of the mind. *McGrawHill, New York*.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. *Psychological bulletin*, *128*(1), 3.

Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. (2014). Large-scale psychological differences within China explained by rice versus wheat agriculture. *Science*, *344*(6184), 603-608.

Triandis, H. C., & Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of personality and social psychology*, *74*(1), 118.

Twenge, J. M., Campbell, W. K., & Gentile, B. (2012). Increases in individualistic words and phrases in American books, 1960–2008. *PloS one*, *7*(7), e40181.

Xu, Y., & Hamamura, T. (2014). Folk beliefs of cultural changes in China. *Frontiers in psychology*, *5*, 1066.

Zeng, R., & Greenfield, P. M. (2015). Cultural evolution over the last 40 years in China: Using the Google Ngram Viewer to study implications of social and political change for cultural values. *International Journal of Psychology*, *50*(1), 47-55.

[包寒吴霜，吴胜涛，郑昊，陆海蓉，郑伟，戴炜，陆梭, & 朱廷劭. (2018). 雾霾下的自我:空气污染与个体主义-集体主义的大数据分析. *社区心理学研究*, 6(2), 71-85.]

[苏红, & 任孝鹏. (2014). 个体主义的地区差异和代际变迁. *心理科学进展*, *22*(6), 1006-1015.]

[王燕. (2016). *互动对 MMORPG 类网络游戏商品消费意愿的影响研究* (Doctoral dissertation，南京：南京邮电大学).]

Use individualism-collectivism words in Weibo to predict players' preference for single player game or online game

ZHANG Weibin[1, 3]; WANG Jiazhou[2, 3]; SA Qieer[2, 3]; WANG Xinrui[2, 3]; DAI Qilong[2, 3]; ZHU Tingshao[1, 3]

([1] CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing 100101, China)

([2] CAS Key Laboratory of Mental Health, Institute of Psychology, Beijing 100101, China))

([3] Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Players' preferences for different types of games are influenced by their own characteristics. The number of players determines that the mode of single player game is more independent, while the mode of online game mode is more collaborative. Given that individualism individuals tend to emphasize independence, collectivism individuals emphasize collaboration. We hypothesized that players' individualism-collectivism tendency may affect their preference for single player game or online game. This study used Weibo user's data to explore whether there was a difference in individualism-collectivism words expressions between single player game players and online game players. Then we used these features to predict players' single player or online game preferences. The result showed that single player game players expressed more individualism words in Weibo, while online game players expressed more collectivism words. Using machine learning method, individualism-collectivism words expressions could predict

players type, but accuracy of the model was low. This study provided preliminary evidence for using Weibo data to identify users' preference for games, thus had certain application value.

Keywords: individualism-collectivism; singer player game; online game; player type; content analysis